

A FRAMEWORK FOR THE OPTIMIZATION OF DISCRETE-EVENT SIMULATION MODELS

B. D. Joshi[#], R. Unal[#], N. H. White^{*} and W. D. Morris^{*}

[#] Dept. of Engineering Management
Old Dominion University
Norfolk, VA 23529

^{*} Vehicle Analysis Branch
NASA Langley Research Center
Hampton, VA 23681

Presented at the 17th ASEM National Conference
October 10-12, 1996
Dallas, Texas

Abstract

With the growing use of computer modeling and simulation, in all aspects of engineering, the scope of traditional optimization has to be extended to include simulation models. Some unique aspects have to be addressed while optimizing via stochastic simulation models. The optimization procedure has to explicitly account for the randomness inherent in the stochastic measures predicted by the model. This paper outlines a general-purpose framework for optimization of terminating discrete-event simulation models. The methodology combines a chance constraint approach for problem formulation, together with standard statistical estimation and analyses techniques. The applicability of the optimization framework is illustrated by minimizing the operation and support resources of a launch vehicle, through a simulation model.

Introduction

Simulations are used to study physical systems or processes that are too complex to permit analytical model formulation and evaluation. The complexity in the physical system is usually attributed to its size, interaction between its subsystems or components, and randomness inherent in its processes. In simulation modeling, the physical system under consideration is mimicked

or imitated with the help of a computer program. The performance or the response of the system can be observed by running the simulation program with operational data. These operating parameters can be changed to obtain a desired improvement in the simulation response and hence performance of the system.

The power of modern computing and the emergence of simulation software in recent years, has contributed to the increasing use of simulation in industry. Simulation models now span diverse application areas such as: designing and analyzing manufacturing systems, characterization of semiconductor devices, determining ordering policies for an inventory system, and analyzing a financial or economic system, to name a few (Law and Kelton, 1991).

With the growing incidence of simulation modeling, it is essential to extend the scope of traditional optimization to include the simulation domain. Optimization procedures for simulation have to specifically account for the randomness inherent in the stochastic simulation models. Often in practice, however, the randomness is ignored and simulation models are analyzed in an ad-hoc manner (Law and Kelton, 1991). In this paper, a unifying framework, based on well established procedures in mathematical programming and statistics, is outlined for use with terminating types of discrete-event simulations. Such a methodology has been generally lacking in the simulation opti-

mization literature, which largely focuses on developing new optimization approaches.

The framework employs the chance constraint approach for treating stochastic constraints for the purpose of problem formulation (Charnes and Cooper, 1956). Standard statistical techniques are employed for estimating the stochastic components of the simulation. Optimization techniques such as path or pattern search methods, can then be applied in conjunction with this framework, to optimize the system. The framework provides guidelines for a rigorous study so that resulting inferences can be made with statistical confidence. The methodology is illustrated by minimizing the operation and support resources for a reusable launch vehicle, modeled through a terminating discrete-event simulation. A genetic algorithm is used to perform the numerical optimization.

The Simulation Optimization Problem

In this section, the general stochastic simulation optimization problem is introduced with the help of an example. Consider a terminating simulation modeling the activities of a bank telling system, in an optimization study. The model simulates the essentially stochastic processes of customer arrival, queues, and teller service and idle times, through appropriate probability distributions. Let us suppose that the optimization problem is as follows:

Minimize the average waiting time for bank customers arriving randomly, while ensuring that the average teller idle time does not exceed a certain level (say 20 minutes), and resources required do not exceed an allowable limit (say \$60,000/ year).

The objective function i.e. ‘minimization of the average time that a customer waits for service’, is stochastic in nature and predicted by the simulation. The problem has two constraints. The ‘average teller idle time not to exceed 20 minutes’ constraint is stochastic and predicted by the simu-

lation. The ‘total resources (teller salaries) not to exceed \$60,000’ constraint is deterministic. Salaries are based on the number of tellers assigned to the simulation by the user at run-time. The tellers assigned clearly influence the resources and system performance. Optimization thus requires determining the number of tellers to assign, so as to minimize the average customer waiting time within the constraint space.

A general simulation optimization problem can thus be stated as determining the settings of the input parameters $X = (X_1, X_2, X_3, \dots, X_p) \in S$, over a region $S \subset R^p$, which optimize the expected response $E(F(X))$ (Jacobson and Schruben, 1989). The region $S \subset R^p$ is defined or bounded by the constraints of the form $E(Z(X)) \geq b$, where $Z(X)$ may be based on the stochastic simulation response.

Such problems are known to be hard to solve (Jacobson and Schruben, 1989). The objective function and/or constraints may be predicted by the stochastic simulation and hence the problem lacks a closed form analytical formulation. Calculus based methods and mathematical programming techniques, requiring explicit problem formulation and continuous domains, cannot be applied (Fu, 1994). Non-linear programming approaches such as path and pattern search methods are generally used to optimize simulated systems. Recently, probabilistic search strategies such as genetic algorithms and simulated annealing have also been used with good results.

Assumptions

Due to its random nature, each run of a simulation has the potential to give rise to slightly different output responses. Hence, the responses predicted by the simulation are generally estimated through a sample collected by replicating the simulation with different random number seeds. Statistics such as the mean, variance, and interval estimates are used for estimation. These statistics, computed in the standard manner, are based on several assumptions. These are:

1. The stochastic process is covariance stationary.
2. The sample variance is an unbiased estimator of the population variance.
3. The observations are independent and identically distributed.

The above assumptions do not always hold true in simulation studies. For example, covariance stationarity may not rigorously hold for terminating simulations, unless the simulation time span is sufficiently long to warrant stationarity. In terminating simulations, a simulation stops when a natural events signaling the end of the simulation run occurs. (For instance, in the bank teller example, the simulation time span may be defined as 8 a.m. to 5 p.m.). Since we do not necessarily run the simulation until steady state, the underlying joint distributions of the random variables may change over time. Also, in order to generate independent and identically distributed observations, a true random number generator is required. However, in practice, pseudo-random number generators are used. Furthermore, Law and Kelton (1991) observe that simulation output data is usually correlated.

Therefore, the above assumptions do not hold in the strictest sense and may be violated by a varying degree. However, there are no other accepted analyses methods for simulation data. Hence it is recommended that regardless of slight violations, standard statistical estimation be used (Law and Kelton, 1991; Pritsker, 1984; Kleijnen, 1987; and Fishman, 1978).

Accuracy

One of the issues associated with a stochastic simulation, is the accuracy with which a stochastic variable is to be estimated. The desired accuracy for each parameter can be specified by the decision maker in terms of statistical confidence intervals. (Confidence intervals state the probability $(1-\alpha)$, that the true mean is actually contained in an interval of width $(\pm w)$, about the estimated mean).

When a simulation involves multiple stochastic variables, the *overall* confidence $(1-\alpha)$ associated

with an optimization study, is based on satisfying the individual confidence intervals $(1-\alpha_i)$ *simultaneously*. Thus the overall confidence satisfies the Bonferroni inequality given by

$$P \geq 1 - \sum \alpha_i,$$

and implies a lower overall accuracy.

For studies involving ten or less stochastic variables, if an overall confidence for the study $(1-\alpha)$ is desired, then the individual confidences $(1-\alpha_i)$ can be selected by the following relation:

$$\sum \alpha_i = \alpha.$$

However, for more than ten stochastic variables, the accuracy of individual variables obtained by the Bonferroni inequality may be prohibitively high. For example, if an overall confidence of 90% is desired for a simulation involving ten variables, the individual confidences have to be at least 99%. (Similarly, if the individual confidence intervals of ten stochastic variables is 90%, the overall confidence is only greater than or equal to zero). Therefore, for such cases, standard 90% or 95% individual confidence intervals are recommended (Law and Kelton, 1991). However, the analysis results in such cases should be interpreted with caution, as one or more of the individual confidence intervals may not contain the corresponding true mean.

Replications

Once the accuracy for each stochastic variable is established, the number of replications required for the optimization study can be determined. The accuracy depends on the sample size or the number of replications used in a study. A large sample size implies that the estimated mean is closer to the true mean, and hence increases the accuracy. A high level of accuracy is usually desired so that the results of the simulation study and hence subsequent decisions can be made with a satisfactory level of confidence. However, due to the finite resources (CPU time) available for a simulation study, the

number of replications possible are usually limited.

The individual confidence levels established above can be used to obtain the required sample size for a desired accuracy in estimating the individual means. The following steps are undertaken to compute the sample size (Law and Kelton, 1991; Kleijnen, 1987):

1. In an initial pilot experiment, the simulation model is run for say 1000 or more replications, to obtain a representative distribution for each stochastic variable.
2. The estimated mean and variance for each stochastic parameter are computed from the pilot experiment.
3. Based on these estimates, the sample size or the required number of replications can be computed for each stochastic variable. For the specified confidence or probability $(1-\alpha)$ that an interval (width $\pm w$ about the estimated mean) contains the true mean, the number of replications is determined as follows:

$$(z_{1-\alpha} s/w)^2$$

where $z_{1-\alpha}$ represents the standard normal statistic covering an area $(1-\alpha)$, and s is the standard deviation of the sample.

The highest number of replications so computed is selected for the simulation study. The individual confidence of variables can be assessed at this new sample size. The variables that required lower replications, will have correspondingly higher confidences at the newly established sample size.

Chance Constraints

The variability inherent in stochastic constraints complicates the simulation optimization problem by forming fuzzy boundaries for the feasible region. This presents a danger of erroneously ac-

cepting a solution as feasible, while it may have a high probability of being infeasible, and vice versa. The chance constraint theory approach can be used to convert the stochastic constraints into deterministic constraints. Chance constraints (Charnes and Cooper, 1959) permit constraint violation up to a pre-specified probability limit. The decision maker expresses a risk tolerance, in terms of a permissible probability of constraint violation. Consider a stochastic constraint of the form $A(x) \leq b$, where $A(x)$ is a simulation response. Using the chance constraint approach, this can be reformulated in terms of risk tolerance as $P(A(x) > b) \leq \alpha$, where α denotes the extent to which constraint violation is permitted.

The chance constraints can be implemented through confidence interval estimates (Teleb and Azadivar, 1994). We know that the confidence associated with an interval estimate denotes the probability that the interval contains the true parameter. For example, the upper limit associated with an interval of confidence $(1-\alpha)$ states that the probability that the true mean exceeds this limit is at most α . Thus, the upper and lower limits of the interval at the specified confidence $(1-\alpha)$ (or risk α) provide deterministic boundaries for the infeasible region. In this manner, confidence intervals provide a means of implementing chance constraints.

Using confidence intervals, the upper limit at confidence $(1-\alpha)$ can be used to denote the chance constraint $P(A(x) > b) \leq \alpha$, as $A(x)_{\text{upp lim}, \alpha} \leq b$. The confidence intervals are estimated by using the Student's t distribution in the standard manner:

$$\begin{aligned} \text{Upp_Lim} &= \bar{x} + \frac{(t_{(n-1), \alpha} s)}{\sqrt{n}} \\ \text{Low_Lim} &= \bar{x} - \frac{(t_{(n-1), \alpha} s)}{\sqrt{n}} \end{aligned}$$

where n is the number of replications, \bar{x} is the estimated mean, s is the standard deviation, and $t_{(n-1), \alpha}$ is obtained from the Student t distribution at $(n-1)$ degrees of freedom and α coverage.

The above confidence intervals based on Student's t are robust to minor deviations from normality of the distribution. However, in cases of serious non-normality and very small sample size, the Johnson's modified t statistic for non-normal distributions, is recommended (Johnson, 1978; Kleijnen, 1987). This adjusted statistic approximates Student's t distribution, by accounting for the skewness of the distribution, thus permitting its use for hypothesis testing and confidence intervals. Johnson's modified t statistic has been used successfully on distributions with varying degrees of non-normality, including the exponential distribution (Johnson, 1978; Kleijnen, 1985). The confidence intervals by the modified statistic are given by

$$\text{Upp_Lim} = \bar{x} + \frac{(t_{(n-1), \alpha} s)}{\sqrt{n}} + \frac{\mu_3}{6s^2 n}$$

$$\text{Low_Lim} = \bar{x} - \frac{(t_{(n-1), \alpha} s)}{\sqrt{n}} + \frac{\mu_3}{6s^2 n}$$

where μ_3 is the third central moment estimated in the standard manner.

Application to Operations and Support Modeling

The above simulation optimization framework is applied to a NASA model simulating the operation and support activities of launch vehicles during conceptual design (Ebeling and Donohue, 1994). It uses estimated values for component reliability and maintainability, to simulate the missions, and pre- and post-flight maintenance. Given the operation and support resources (individual crew assigned to nine maintenance sub-systems c_n , and vehicles in the fleet v) the model predicts the number of missions flown and the mean vehicle turn-time for a particular space program. The problem is to:

Minimize the cost function

$$100*v + c_1 + c_2 + c_3 + c_4 + c_5 + c_6 + c_7 + c_8 + c_9$$

subject to stochastic constraints evaluated by the simulation:

- (1) Average delay $E(\text{delay})$ not to exceed 48 hours
- (2) Average mission rate of the space program $E(\text{suc_mis})$ is met (140 missions in five years).

Minimization of the resources thus requires the determination of the smallest fleet and maintenance crew size, that enables meeting the target mission rate in a timely manner.

Problem Formulation The stochastic constraints were converted to deterministic constraints by using the chance constraint approach. The NASA decision makers expressed a risk tolerance, in terms of permissible probability of constraint violation, of 5%. These risk tolerances are stated symbolically as:

$$P[E(\text{delay}) > 48 \text{ hours}] \leq 0.05$$

$$P[E(\text{suc_mis}) < 140] \leq 0.05$$

The limits at the specified 95% confidence (5% risk tolerance) were established by using the Student's t distribution as described above. These limits provide deterministic boundaries for the infeasible region, as follows:

$$\text{delay}_{\text{upp_lim}, .05} \leq 48 \text{ hours}$$

$$\text{suc_mis}_{\text{low_lim}, .05} \geq 140$$

Accuracy and Sample Size The accuracy desired was specified in terms of confidence $(1-\alpha)$ that the true mean is within the interval $\pm w$, by NASA engineers as:

	Desired width $\pm w$	Desired conf. $(1-\alpha)$
Delay	± 48 hours	80%
Successful Missions	± 2	95%

The number of replications required were computed using the steps outlined in the preceding

sections. This computation yielded: 19 replications for the average delay, and 18 replications for successful missions. Hence, a conservative sample size of 20 was selected.

Results The problem formulated above was solved by applying a genetic algorithm (Gage, 1995). A genetic algorithm is a biased random search approach that has been used successfully for optimization in complex landscapes (Holland 1975; Goldberg, 1989). Unlike other mathematical programming techniques, a genetic algorithm does not require continuity, differentiability, and explicit problem formulation. Due to these reasons, the genetic algorithm is an attractive technique for the simulation domain.

Inspired by the mechanics of natural selection and evolution, a genetic algorithm uses binary strings to encode parameters, for representing solutions in the search space. Based on the fitness of such binary strings, contained in a randomly selected initial generation, subsequent generations are created by mating and mutation of the strings. The genetic algorithm continues to form new generations, in this manner, until a suitable criterion is reached. For a detailed description of genetic algorithms please see Goldberg (1989), and Srinivas and Patnaik (1994).

The minimum objective function found by the genetic algorithm when applied to the problem formulated above was 273. Two vehicles and 73 crew contribute to this objective function. The constraints at this input combination were satisfied and well within their tolerance levels. The mean of average delay was 15.73 hours, with a 95% confidence that the delay does not exceed 38 hours. The target mission rate of 140 missions in a five year time span was achieved with a 95% confidence. The above stochastic measures were estimated at an 80% confidence of being within ± 2 days, and 95% confidence of being within ± 2 missions, respectively. The overall accuracy of the optimization study was lower at 75% confidence.

Conclusions

Stochastic simulation modeling presents some unique problems that have to be addressed during optimization. In this paper a general purpose framework has been outlined for the optimization of terminating discrete event simulations. A unifying framework or methodology for treating the stochastic components of such problems has been generally lacking in the literature. The methodology presented here employs a chance constraint approach to convert the fuzzy boundaries presented by stochastic constraints into deterministic constraints. Standard statistical techniques are used to perform the optimization study at desired levels of confidence. Therefore the optimization framework is based on a sound statistical background. Inferences drawn from such studies can be stated in terms of interval estimates, which are more robust than point estimates. This provides more information while making decisions based on the simulation study. The practical significance of the methodology was illustrated by applying it to a NASA simulation. The operation and support resources of reusable launch vehicles were successfully minimized with the help of a genetic algorithm.

Acknowledgement

This study was supported in part by NASA contract NAS1-19858-13.

References

- Charnes A. and W. W. Cooper, "Chance Constrained Programming" *Management Science*, Vol. 6, No. 1, 1959.
- Ebeling, C. E., and C.B. Donohue, Integrating O&S Models During Conceptual Design - Part III, Annual Report for NASA Langley, Grant No. NAG1-1-1327, 1994.
- Fishman, G.S., Principles of Discrete Event Simulation, John Wiley, New York, 1978.

Fu, M.C., "Optimization Via Simulation: A Review", *Annals of Operations Research*, Vol. 53, pp 199-247, 1994.

Gage, P. J., *New Approaches to Optimization in Aerospace Conceptual Design*", NASA Report 196695, Contract Number NAG2-640, 1995.

Goldberg, D., Genetic Algorithms in Search, Optimization, and Machine Learning, Addison-Wesley Publishing Company, Reading, 1989.

Holland, J., Adaptation in Natural and Artificial Systems. The University of Michigan Press. Ann Arbor, 1975

Jacobson, S.H., and L.W. Schruben, "Techniques for Simulation Response Optimization", *Operation Research Letters*, Vol. 8, pp 1-9, 1989.

Johnson, M., "Modified *t* Tests and Confidence Intervals for Asymmetrical Populations", *Journal of the American Statistical Association*, Vol. 73, pg 536-544, 1978.

Kleijnen, J., G. Kloppenburg and F. Meeuwsen, "Testing the Means of an Asymmetric Population: Johnson's Modified *t* Test Revisited", *Communications in Statistics, Simulation and Computation*, Vol. 15, No.3, pp 715-732, 1986.

Kleijnen, J., Statistical Tools for Simulation Practitioners, Marcel Dekker, New York, 1987.

Law, A., and W. D. Kelton, Simulation Modeling and Analysis, Mc-Graw Hill, New York, 1991.

Pritsker, A.A.B., Introduction to Simulation and SLAM II, Systems Publishing, West Lafayette, IN, 1984.

Srinivas, M and L. Patnaik, "Genetic Algorithms: A Survey", *Computer*, vol. 27, no.6, 17 pgs. 1994.

Teleb, R., and F. Azadivar, "A Methodology for solving multi-objective simulation-optimization problems", *European Journal of Operational Research*, Vol. 72, pp 135-145, 1994.

B. D. Joshi is a doctoral student in the department of Engineering Management at Old Dominion University. Her research interests are in the areas of simulation, optimization, and robust engineering. She is a member of IEEE and INFORMS.

R. Unal is an associate professor of Engineering Management at Old Dominion University. His research and teaching interests are in the areas of multidisciplinary design optimization, response surface methods, design of experiments, and robust design for cost. He is a member of American Society for Engineering Management, American Society for Quality Control, and International Society of Parametric Analysis.

N. H. White is a computer engineer in the Vehicle Analysis Branch, NASA Langley Research Center.

W. D. Morris is an aerospace engineer in the Vehicle Analysis Branch, NASA Langley Research Center.